



Zhang, N., Wang, W., Harris, R., & Leckie, G. (2020). Analysing inter-provincial urban migration flows in China: A new multilevel gravity model approach: A new multilevel gravity model approach . *Migration Studies*, 8(1), 19–42. [mny026].  
<https://doi.org/10.1093/migration/mny026>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1093/migration/mny026](https://doi.org/10.1093/migration/mny026)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Oxford University Press at <https://doi.org/10.1093/migration/mny026> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Analysing inter-provincial urban migration flows in China: A new multilevel gravity model approach

Xingna Nina Zhang<sup>†,‡,\*</sup>, Wenfei Winnie Wang<sup>†</sup>, Richard Harris<sup>†</sup> and George Leckie<sup>‡</sup>

<sup>†</sup>School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK and

<sup>‡</sup>Centre for Multilevel Modelling, Graduate School of Education, University of Bristol, Bristol, BS8 1SS, UK

\*Corresponding author: Email: [nina.zhang@bristol.ac.uk](mailto:nina.zhang@bristol.ac.uk)

## Abstract

This paper proposes a novel multilevel gravity model of migration to study the under-researched topic of urban to urban migration in China. Many previous studies have looked at rural to urban migration in the context of urbanisation and economic development, and at return migration. Very few have looked at what is becoming more important in increasingly urbanised countries, which is the movement from one urban location to another. In the study, we develop a new method that allows for the inter-connections between migration flows: between those that share an origin, those that share a destination, and where there is a reciprocal flow between places. A conventional gravity model of migration ignores those connections, risking erroneous estimation of the regression parameters and of their statistical significance. It also ignores that those connections are of substantive interest—they reveal the interconnections between places regarding the numbers of migrants that they send and receive. We motivate and illustrate the advantages of our approach using 2010 interprovincial migration census data for China. The results obtained from the model confirm the effect of distance, of population size and of regional income levels. They show that there is greater variation in the numbers of migrants received by provinces than there is in the numbers sent, and that reciprocal migration between pairs of provinces is an important feature of what is happening in China, especially between the neighbouring provinces of Sichuan and Tibet.

**Keywords:** migration, urban, China, gravity model, multilevel model

doi:10.1093/migration/mny026

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

# 1. Introduction

This paper proposes a new method to study between province urban migration flows in China. Migration is a topic of enduring interest in population studies (De Haas 2010a,b; Molho 2013). The literature shows that migration is a complex social-economic phenomenon exhibiting different features in different geographical contexts. Developed economies recently have seen a rise in counter-urbanisation population movement (Remoundou, Gkartzios and Garrod 2016), while migration from rural to urban areas has been the predominant trend in the developing world (De Haas 2010a,b). The focus on rural-to-urban and urban-to-rural migration (Ezra and Kiros 2001; Fan and Wang 2008; Remoundou Gkartzios, and Garrod 2016) relates to the classic two-sector migration theory whereby labour transfers from the primary and rural employment sectors to the secondary and urban sectors at the beginning stage of industrialisation (Lewis 1954; Harris and Todaro 1970; Todaro Michael 1976), and also to the post-industrialisation migration framework, which emphasises urban-to-rural movement and population decentralisation processes away from cities into less densely populated areas (Berry 1976; Vartiainen 1989).

Migration between urban areas (thereafter urban–urban migration) is less studied perhaps because it fits into neither framework. It is, however, prevalent in both developed and developing worlds. In industrialised countries where urbanisation is reaching its saturation level, the majority of movements are between urban areas. Urban–urban migration has been studied as part of the dynamics of urban systems (Andersson et al. 2012) or as a component contributing to differential urban growth (Pumain and Sanders 2013). In addition, urban–urban migration has been studied as a spatial movement between different labour markets most closely linked with economic factors such as employment and income (Flowerdew and Salt 1979; Poot 1986) but also with amenities (Greenwood and Hunt 1989) and house prices (Johnston et al. 2016).

In the developing world, while rural–urban migration is the dominant trend, urban–urban migration has become the main form of population movement in Latin America since the 1980s due to its accelerated urbanisation process (Cerrutti and Bertonecello 2003). Generally, however, there have been limited attempts to examine how the pattern and process of urban–urban migration vary from rural–urban migration (Machado and Hakkert 1988; Shefer and Steinvortz 1993). Indeed, little is known about what may be the world's largest urban–urban migration, which is occurring in China and is caused by rapid urbanisation with lessening institutional restrictions and rising population mobility (Feng et al. 2002; Wu and Yao 2003). There has been some discussion of how urban–urban migrants integrate with mainstream society and access the welfare system in a few cities (Yang 2013; Cheng, Nielsen, and Smyth 2014), and on the links between increasing urbanisation, growing urban–urban migration and the career-driven characteristics of urban–urban migrants (Vignoli 2008; Hahn 2010). Nevertheless, broader understanding of the macro-level urban–urban migration patterns and its mechanisms are unclear. This is despite China's census reporting 260 million people to have migrated internally in 2010 (approximately 20 per cent of the total population), a third of which were urban–urban migrants.

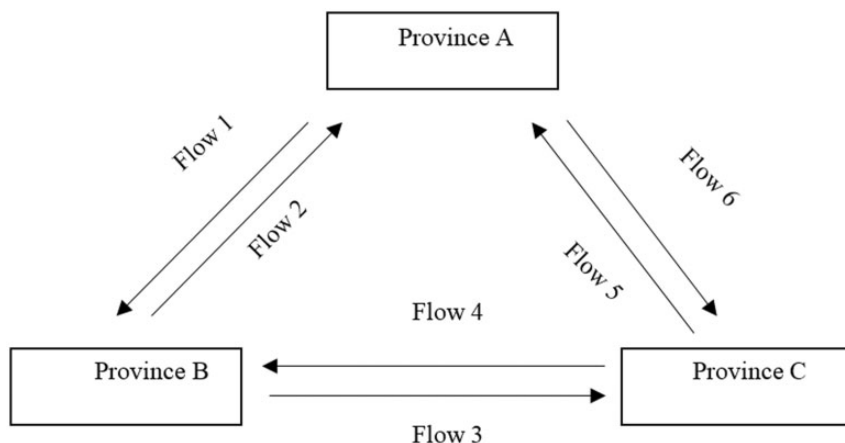
This paper addresses two gaps in the literature: the lack of attention given to urban–urban migration in China, and the lack of an appropriate statistical model to do so. The study proposes a multilevel gravity model of migration which combines the merits of the linear regression formulation with multilevel modelling to investigate: (a) how flows originating from the same Chinese province and (b) how flows ending at the same province vary from each other in regard to the average number of migrants they contain; (c) what is the correlation between the average out-migration and in-migration flows across provinces; and (d) how the reciprocal flows between two provinces are related. We illustrate the model using data about urban–urban migration flows within China obtained from the 2010 Chinese census.

## 2. The gravity model, multilevel modelling and inter-flow dependencies

The gravity model is widely used in analysing migration flows (Fan 2005), where the numbers of people moving between locations are modelled as a function of the attributes of the locations such as population size and GDP, and of the physical or socio-economic distance between places (Converse 1949; Christian and Braden 1966).

There are several reasons for its popularity; amongst them are that the gravity model is capable of incorporating both origin and destination attributes when modelling migration flows (Beine, Bertoli, and Moraga 2014). The model also is flexible in allowing predictors of migration to be added beyond the original form of the model where only distance and populations are used (Beine, Bertoli, and Moraga 2014; Shen 2015). In China's context, among the existing gravity model studies, some have been conducted to determine how the total migration flows distribute across space and over time (Fan 2005), and others focus on the determinants of the total migration flows (Shen 2012).

Conventionally gravity models are formulated and estimated as linear regression models, where each row of the data matrix represents a tally of movements from one place to another—a flow—and these flows are tacitly assumed to be independent of one another. However, Origins and destinations can be connected in four key ways as illustrated in Fig. 1: Type 1, indirectly when multiple destinations receive migrants from the same origin (e.g. flows 1 and 6); Type 2, when multiple origins send migrants to the same destination (e.g. flows 2 and 5); Type 3, when an origin sends to a destination, which is itself a destination to another origin (e.g. flows 3 and 1); and Type 4, directly when there is migration in both directions between two places, so both places are simultaneously an origin and a destination to the other (e.g. flows 1 and 2). These connections suggest that the flows between places are not independent of one another. However, these four flow dependencies are seldom addressed in the regional migration literature. If the assumption of independence is invalid and there are dependencies between the flows then the estimates of statistical significance and of effect size are affected (the former typically are over-stated, whereas the latter may have deviated from their true value). In fact, some degree of dependency is almost inevitable: for a gravity model, each row in the data matrix provides an origin, a destination and the number of people that moved between them, as well as other



**Figure 1.** Illustration of the different potential population flows between three provinces and how they are interrelated.

attributes of the places that may explain the flow. Unless those origins and destinations are all unique, then those attributes are necessarily repeated, creating group dependencies which ought to be controlled for.

Aside from the potential for estimation errors, standard linear regression lacks the capacity to quantify the strength of the inter-dependencies between the flows. Multilevel modelling can be seen as a generalisation of linear regression, designed to deal with hierarchically structured or cross-classified data (Goldstein 2011; Leckie 2013; Harris 2017), which has the potential to estimate similarities between observations that belong to a common group—a common origin, for example. Only a small number of migration studies have used multilevel modelling techniques. Amongst those that have, some examined the determinants of migration among certain migrant groups at different geographical scales (Kallan 1993; Ezra and Kiros 2001), whilst others investigated interregional flows but treating in-migration and out-migration as independent events (Dennett and Wilson 2013). In the context of China, Yang and Guo (1999) use multilevel modelling to examine gender differences in the determinants of labour migration at the individual/household and community levels. While Shen (2016) uses a two-stage Poisson version of the gravity model with a network spatial filter and decomposes the estimation errors into the overall effect of the constant, of the relative emissivity of the origin and the relative attractiveness of the destination, and a measure of interaction between places.

What have not been systematically examined are the potential connections between origins and destinations—the ways that the flows are related to and dependent upon one another. A partial exception was undertaken by Thomas, Stillwell and Gould (2015), who used multilevel modelling to allow for individual and contextual variations by origin and by destination in the distances moved by residential migrants in England and Wales, but did not consider the correlations between origins and destinations in terms of the migration flow.

In this paper, we are interested in allowing for and quantifying the four types of flow dependency which arise when two flows share provinces in common. First, two flows may correlate due to sharing a common origin (Type 1). Second, two flows may correlate due to sharing a common destination (Type 2). Third, two flows may correlate if the destination in the first flow is the origin in the second flow (Type 3). Fourth, two flows may correlate if there are reciprocal flows between two places (Type 4). These dependencies are in part induced by unmodeled origin and destination effects. Origin effects reflect variation between places in the number of migrants that move away. Destination effects reflect variation between places in terms of the volume of migrants they attract. These two sets of effects may well be associated, reflecting a correlation between the in-migration and out-migration flows for a place when it both sends out and receives migrants. Even if we take into account these correlated origin and destination effects the residuals may well continue to be correlated within pairings of provinces, reflecting the bilateral flows between themselves. All these four effects represent the spatial dependencies between places—how the numbers of migrants sent or received at one place can be dependent on the number of migrants sent or received at others. Instead of treating space as a vacuum in which all that matters is the mass of attraction between any two places that are somehow isolated from other places, we instead adopt a system-wide perspective and allow for the inter-connections between places through migration flows. Although the importance of spatial dependency has been stressed in multiple fields of social science (Fingleton 1986; Getis 1990; Leorato and Mezzetti 2016), it has not received comparable attention in migration studies. To fill this empirical gap, we propose multilevel modelling extensions to the traditional linear regression formulation of the gravity model of migration.

### 3. Methodology

We start this section by introducing the gravity model of migration. We then present three increasingly realistic implementations of this model: Model 1, the traditional linear regression formulation of the gravity model; Model 2, a standard cross-classified multilevel model formulation which extends Model 1 to capture systematic variation in out- and in-migration across provinces; and Model 3, an extended version of Model 2 where we additionally allow for correlations in the out and in-migration flows. We then show how only Model 3 captures the four dependencies that arise in migration data. Finally, we discuss estimation and the illustrative data.

#### 3.1 The gravity model of migration

Let  $m_{ij}$  denote the number of migrants who move from origin province  $i$  ( $i = 1, \dots, n$ ) to destination province  $j$  ( $j = 1, \dots, n$ ). The traditional gravity model of migration can then be written as

$$m_{ij} = k \frac{p_i^a p_j^b}{d_{ij}^c} \quad (1)$$

where  $p_i$  and  $p_j$  denote the populations of province  $i$  and  $j$ ,  $d_{ij}$  denotes the distance between them,  $k$  is a constant, and  $a$ ,  $b$  and  $c$  are the powers originally hypothesized to take the values 1, 1 and 2, respectively, resembling Newton's law of gravity (Christian and Braden 1966; Claeson 1969). The underlying assumption is that a migration flow between two places is determined by the attraction between the origin and the destination, which can be measured as proportional to the product of the 'masses' of two places (the origin and destination populations in this case) and inversely proportional to the square of the distance between them. However, this assumes that  $m_{ij} = m_{ji}$  whereas, in practice, the bilateral migration flows between pairs of places are often not equal to each other in the number of migrants that they contain. Furthermore, there is no reason why 'mass' should be measured only by population size. Thus the original form of the gravity model was soon refined and the artificial constraints on power coefficients of  $a$ ,  $b$  and  $c$  were then released (Ginsberg 1972; Wilson 1971).

### 3.2 Model 1: the traditional linear regression formulation

Taking the natural logarithm of Equation 1 and adding a residual  $\varepsilon_{ij}$  results in the following log-normal linear regression formulation of the gravity model where  $a$ ,  $b$  and  $c$  are now regression coefficients to be estimated from the data

$$\ln(m_{ij}) = \ln(k) + a \ln(p_i) + b \ln(p_j) - c \ln(d_{ij}) + \varepsilon_{ij}; \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2). \quad (2)$$

The log-log form of this model leads these parameters to be interpreted as elasticities (Leamer 2012); the relative change in the conditional expectation of  $m_{ij}$  associated with a unit relative change in the relevant covariate. For example, regression coefficient  $a$  is approximately equal to the percent increase in  $m_{ij}$  associated with a one percent increase in  $p_i$  holding all other covariates constant.

Our first model of interest, Model 1, is a natural extension of Equation 2 to include further predictors of migration. It is convenient to write this more general model as follows

$$y_{ij} = \beta_0 + \mathbf{x}'_{1i} \beta_1 + \mathbf{x}'_{2j} \beta_2 + \mathbf{x}'_{3ij} \beta_3 + \varepsilon_{ij}; \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad (3)$$

where  $y_{ij}$  denotes the log migration flows,  $\mathbf{x}_{1i}$  is a vector of province-level covariates including log population and other demographic, social, environmental and/or economic attributes of origin  $i$ ,  $\mathbf{x}_{2j}$  is an equivalent vector of province-level covariates for destination  $j$ ,  $\mathbf{x}_{3ij}$  is a vector of province pair-level covariates including the log of the distance between them, and  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are the corresponding coefficient vectors.

Flowerdew and Aitkin (1982) proposed a Poisson regression version of Equation 3 which has various advantages over the linear regression formulation and can become appreciable when migration flows are low. Our application, however, consists of very large migration flows and so it is the linear rather than Poisson formulation of the gravity model that we extend here.

### 3.3 Model 2: the standard cross-classified multilevel formulation

A fundamental limitation of the linear regression model is that it assumes the residual migration flows  $\varepsilon_{ij}$  are independent. However, we expect residual migration flows to systematically vary across origins and destinations. Specifically, we expect the out-migrations from a given province to be positively correlated as they share a common origin. Likewise, we expect the in-migrations to a given province to be positively correlated as they share a common destination. Linear regression ignores these dependencies and will therefore estimate spuriously precise regression coefficients raising the risk of type I errors of inference: we might conclude covariates to be significant when they are not. The estimated parameters may also be ‘unstable’ and deviate from their true values.

We propose a multilevel modelling based approach to dealing with the complex residual dependencies which arise when modelling migration flows. However, before we introduce this approach, we note that there are other general approaches to dealing with simpler forms of clustered data which could potentially also be applied to migration data. In particular, one might attempt to account for the Type 1 and 2 dependencies introduced above by replacing the model-based standard errors in the usual linear regression formulation of the gravity model with their two-way cluster-robust counterparts (Cameron, Gelbach and Miller 2011). However, this approach does not straightforwardly allow for the more nuanced Type 3 and Type 4 dependencies we aim to capture. Furthermore, the two-way cluster-robust standard error approach does not additionally quantify and therefore allow one to substantively interpret the magnitudes of these four forms of dependencies, nor does it allow one to make predictions regarding specific province origin and destination effects. A central argument in this article is that both these lines of investigation are substantively insightful when studying migration flows.

We address this concern by specifying a multilevel version of the model which includes cross-classified origin and destination random effects to account for systematic residual variation in out-migration and in-migration across provinces. Model 2 can be written as

$$\begin{aligned} y_{ij} &= \beta_0 + \mathbf{x}'_{1i}\beta_1 + \mathbf{x}'_{2j}\beta_2 + \mathbf{x}'_{3ij}\beta_3 + o_i + d_j + e_{ij} \\ o_i &\sim N(0, \sigma_o^2) \\ d_j &\sim N(0, \sigma_d^2) \\ e_{ij} &\sim N(0, \sigma_e^2) \end{aligned} \quad (4)$$

where  $o_i$  and  $d_j$  denote the origin and destination random effects and  $e_{ij}$  the revised residual.

The random effects and residuals are typically stated to be normally distributed with zero means and constant variances. However, we note that normality of the random effects and residuals are not required for consistent estimation of the model parameters and standard errors. It should be kept in mind that empirical Bayes predictions of the random effects do rely on at least approximate normality and so we recommend that one checks this assumption when we apply these models to the data. The origin and destination variances  $\sigma_o^2$  and  $\sigma_d^2$  quantify the degree to which origins and destinations vary in average out-migration and average in-migration having adjusted for the covariates. The residual variance  $\sigma_e^2$  quantifies the remaining variation.



Dividing each variance component by the total residual variance  $\sigma_o^2 + \sigma_d^2 + \sigma_e^2$  gives variance partition coefficients (VPCs) which can be used to quantify the relative importance of origins and destinations in explaining residual migration. For instance, [Thomas, Stillwell and Gould \(2015\)](#) used VPCs to estimate the distances moved by residential migrants in England and Wales and found that city-region random effects are more important than the neighbourhood random effects. The VPCs allow explanation of whether there are unexplained differences between provinces in terms of the numbers of migrants they send or receive or whether the differences between what occurs and what the model predicts is simply random between the individual flows with no evidence of place effects.

### 3.4 Model 3: the extended cross-classified multilevel formulation

Model 2 is an improvement on the standard regression but still assumes that provinces' origin and destination effects are independent of one another. However, a province's level of out-migration is likely to be linked to their level of in-migration, even after adjusting for the covariates. For example, provinces which in general exhibit higher than expected out-migration might be expected to exhibit lower than expected in-migration and vice versa. Put simply, we might expect variation in net-migration over and above that predicted by the characteristics of the provinces captured by the covariates. In Model 3, we therefore allow the origin and destination random effects to correlate  $\text{Corr}(o_i, d_i) = \rho_{od}$ .

Model 2 also assumes that residual migration flows are independent within each pair of provinces. However, here too we might expect a systematic relationship. Namely, where there is a higher than expected flow from one specific province to another specific province we may see a higher than expected flow in the return direction. That is, it seems likely that we might see particular province pairings which exhibit higher (or lower) than expected migration flows in both directions. In Model 3, we therefore also allow for correlated within province-pair residuals,  $\text{Corr}(e_{ij}, e_{ji}) = \rho_{ee}$ .

Model 3 can therefore be written as

$$y_{ij} = \beta_0 + \mathbf{x}'_i \beta_1 + \mathbf{x}'_j \beta_2 + \mathbf{x}'_{3ij} \beta_3 + o_i + d_j + e_{ij}$$

$$\begin{pmatrix} o_i \\ d_i \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_o^2 & \\ & \sigma_d^2 \end{pmatrix} \right\}$$

$$\begin{pmatrix} e_{ij} \\ e_{ji} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & \\ & \sigma_{ee} \end{pmatrix} \right\} \quad (5)$$

where the origin-destination and residual correlations can be derived from the associated variance and covariances parameters in the usual way,  $\rho_{od} = \sigma_{od} / \sigma_o \sigma_d$  and  $\rho_{ee} = \sigma_{ee} / \sigma_e^2$ .

We further specify [Equation \(5\)](#) as  $\mathbf{x}'_i = [\ln(p_i) \ln(I_i)]$  and  $\mathbf{x}'_j = [\ln(p_j) \ln(I_j)]$ , where  $I_i$  and  $I_j$  are the incomes of the origins and destinations, in line with what is commonly used in most studies and reflecting the importance of economic factors in the migration system ([Fan 2005](#); [Beine, Bertoli and Moraga 2014](#)). The origin population  $p_i$  is used to represent the migration potential of the sending place. According to the International Organization for Migration ([Laczko, Tjaden and Auer 2017](#)), migration potential is a

**Table 1.** Equations for the model-implied correlations conditional on the covariates between flows corresponding to the four key dependencies shown in Fig. 1

Dependency	Correlation	Model 1	Model 2	Model 3
Type 1	$\text{Corr}(y_{ij}, y_{ik})$	0	$\frac{\sigma_o^2}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$	$\frac{\sigma_o^2}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$
Type 2	$\text{Corr}(y_{ij}, y_{kj})$	0	$\frac{\sigma_d^2}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$	$\frac{\sigma_d^2}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$
Type 3	$\text{Corr}(y_{ij}, y_{jk})$	0	0	$\frac{\sigma_{od}}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$
Type 4	$\text{Corr}(y_{ij}, y_{ji})$	0	0	$\frac{2\sigma_{od} + \sigma_{ee}}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$

Note. See Appendix for derivations.

valuable indicator of the actual migration flows that take place. The total number of individuals who are ready to migrate is proportional to the size of the population in the sending area. So the larger the origin population, the greater migration potential a place will have. In contrast, the destination population  $p_j$  provides a strong proxy for the employment prospects a place has. According to neoclassical economics (Massey et al. 1993), migrants tend to move to places with more favourable employment conditions. A larger population suggests a bigger and more diverse labour market which provides more and varied employment opportunities. Therefore, it will be much easier for migrants with different specialized skills to find work in destinations with larger populations. This is especially true for speculative migrants (Gordon 1995).

In order to present the multilevel extension of the migration gravity model in as simple an accessible form as possible we choose to enter population into the model linearly. While this has also been the choice made in most past applications of the gravity model, the effect of population on migration may in practice be non-linear. We therefore encourage future research to explore this possibility. One way to do this is to simply enter origin and destination population into the model as polynomial functions (e.g. a quadratic or cubic relationship) rather than as linear terms.

We note that Model 3 takes the same form as the social relations model (Kenny and Kashy 2011), which has recently been adopted to handle counts response variables (Koster and Leckie 2014) and will be of interest to researchers who prefer working with Poisson frameworks (Shen 2016).

Table 1 presents equations for the model-implied correlations between migration flows corresponding to the four key dependencies that we have identified in migration data (see Appendix for derivations). Type 1: two flows may correlate due to sharing a common origin; Type 2: two flows may correlate due to sharing a common destination; Type 3: two flows may correlate if the destination in the first flow is the origin in the second flow; Type 4: two flows may correlate if they are reciprocal flows between the same two places. Model 1 (Equation 3), the linear regression formulation implicitly assumes all these correlations to be zero, and therefore ignores all four dependencies. Model 2 (Equation 4), the standard cross-classified multilevel formulation allows us to estimate the first two correlations and therefore allows for the first two types of dependency. Only Model 3

**Table 2.** The response variable and covariates used for the analysis

Original data	Level	Units	Observations	Mean	SD	Min	Max
Urban–urban migration	Flow	00s persons	930	1,447.80	3,551.79	1	48,921
Urban population	Province	Millions of persons	31	12.40	7.06	0.44	29.58
Urban income	Province	000s yuan	31	18.07	4.70	13.19	31.84
Distance	Flow pair	km	465	1,379.53	729.31	113.69	3,598.79

(Equation 5), our extended cross-classified multilevel formulation with correlated origin and destination effects and correlated within province pair residuals, allows us to estimate all four correlations and therefore allow for all four forms of dependency.

We fit the models by iterative generalised least squares (equivalent to maximum likelihood estimation) using MLwiN 2.36 where we call MLwiN (Rasbash et al. 2009) from within Stata 14 using the user-written runmlwin command (Leckie and Charlton 2013). A script is available from the lead author.

3.5 Data

Data used in this study are mainly drawn from China’s 2010 Census (migration and population data) and 2011’s China Statistical Yearbook (income data). A summary of the data used is given in Table 2. Each variable is log-transformed in the models but is shown in its original scale in Table 2. Origin-destination distance is calculated as the distance between provincial capital cities. Urban income is defined as per capita disposable income of urban households. We choose not to further amend this variable by spatial deflators such as the Consumer Price Index (CPI). Whilst the provincial CPIs could help to justify the absolute price comparisons across localities (Brandt and Holz 2006), there are practical reasons why we do not account for the spatial price difference in the model. Firstly, economic migration decisions are mostly based on relative prices, not absolute prices (Taylor 1999; King 2012). For instance, migrants rarely do an accurate calculation to extensively weigh up the provincial difference of the Consumer Price Index (CPI) in urban areas before moving. Secondly, the deflation of ‘cost-of-living’ is only true if the research subjects buy the same basket as the general local residents with which the local CPI is constructed. This is unlikely in the case of migrants, as migrants usually have distinctive consumption behaviours when compared with local residents (Zukin 1998; Zarate-Hoyos 2004; Giles and Yoo 2007; Han and Chen 2016; Cao et al. 2017; Jiang et al. 2017). In any case, having calculated the urban prices across provinces in 2010 using the calculation procedure of Brandt and Holz (2006), we find little spatial difference due to the similar urban CPIs (Mean = 103.32; S.D. = 0.60) across provinces. Consequently, we do not anticipate that they would have great effect on the migration flows.

**Table 3.** Results from linear regression and multilevel formulations of the gravity model of migration

Variables	Model 1		Model 2		Model 3	
	Est.	SE	Est.	SE	Est.	SE
Fixed part						
Constant	4.995*	0.613	7.534*	1.396	7.410*	1.503
Log of origin urban population	0.921*	0.032	0.909*	0.075	0.909*	0.074
Log of destination urban population	0.632*	0.032	0.619*	0.106	0.620*	0.105
Log of origin urban income	-0.990*	0.122	-1.068*	0.284	-1.064*	0.283
Log of destination urban income	2.182*	0.122	2.104*	0.403	2.108*	0.401
Log of distance	-0.818*	0.044	-1.105*	0.039	-1.091*	0.051
Random part						
Origin province variance			0.106*	0.030	0.104*	0.030
Destination province variance			0.224*	0.060	0.221*	0.060
Individual flow variance	0.641*	0.030	0.348*	0.017	0.348*	0.021
Origin-destination correlation					0.105	0.193
Flow-pair correlation					0.719*	0.023
Dependency (correlation conditional on the covariates)						
Type 1: Common origin			0.156*	0.040	0.155*	0.040
Type 2: Common destination			0.331*	0.062	0.328*	0.062
Type 3: Destination in first flow is origin in the second					0.024	0.044
Type 4: Reciprocal flow (shared origin and destination)					0.419*	0.090
Deviance	2,225.8		1,819.7		1,503.3	

Note: Response variable is the log migration flow (in 1,000s). Est., estimate. SE, standard errors.

\*Denotes  $p < 0.001$ .

The 2010 Census contains a long-form with 10 per cent sampling and a short-form for the whole population. All the independent population variables are from the short-form dataset, while the response variable, the number of urban–urban migrants moving between provinces, is calculated from migration information contained in the long-form dataset. We define interprovincial urban–urban migration as the movement made by migrants with urban Hukou (the household registration system) between urban areas of different provinces, where urban areas includes (a city's) 'Street' and 'Neighbourhood committees of the town'. There are 31 provincial administrative units (thereafter referred to as provinces) in China, generating 930 urban–urban migration streams between them. Those streams are the focus of our study.

## 4. Results

Table 3 shows the results from models 1, 2 and 3: the linear regression, standard cross-classified multilevel and extended cross-classified formulations of the gravity model,

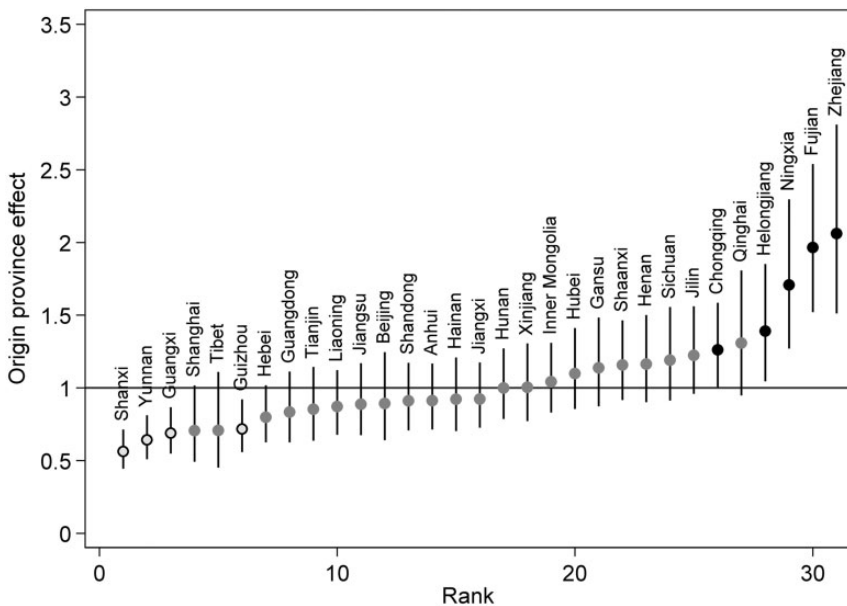
respectively. All three models include the natural log of the origin and the destination populations and incomes as covariates as well as the natural log of the distance between each pair of provinces. Recall that Model 2 extends Model 1 by introducing the origin and destination random effects, while Model 3 further allows for the effects to correlate. Likelihood ratio tests show that Model 3 is significantly preferred to Model 2 ( $\chi^2_2 = 406.1$ ,  $p < 0.001$ ) which in turn is preferred to Model 1 ( $\chi^2_2 = 316.4$ ,  $p < 0.001$ ).

In all three models the estimated coefficients are in the expected directions, similar in magnitude across models, and statistically significant at the 0.1 per cent level (Table 3). The models show that the larger the population of the origin and/or the destination, the greater the flow between them. Specifically, a 10 per cent increase in the origin population is associated with an approximate 9 per cent increase in out-migration, all else being equal, while a 10 per cent increase in the destination population is associated with an approximate 6 per cent increase in in-migration. Origins with lower income send out more migrants and destinations with higher urban income attract stronger migration streams in accordance with the neo-classical economy migration theory (De Haas 2010a,b). Specifically, a 10 per cent decrease in the origin income is associated with an approximate 11 per cent increase in out-migration while a 10 per cent increase in destination population is associated with an approximate 21 per cent increase in in-migration. As expected the distance between the provinces acts as an impeding factor (the longer the distance the weaker the flow).

While the magnitude of the coefficients is similar across models, it is important to note that the standard errors differ dramatically. Moving from Model 1 to Model 2, the standard errors of the province level covariates (incomes and populations) approximately double when we take into account the clustering of migration flows by origins and destinations. The smaller standard errors in Model 1 are therefore spuriously precise, illustrating that the standard linear regression formulation of the gravity model is inadequate for modelling migration flows with shared origins and/or destinations. Moving from Model 2 to Model 3 sees no further change to the standard errors of the province-level covariates; rather it is now the standard error of the flow-pair level covariate (distance) which increases (by approximately 30 per cent) when we additionally take into account the correlation between reciprocal flows. Thus, where interest lies in flow-pair level covariates, even Model 2 the standard cross-classified multilevel model proves insufficient.

For Model 3, the origin, destination and residual VPCs account for 16 per cent, 33 per cent and 52 per cent of the total residual variance, respectively. Thus, having adjusted for the covariates, we see that provinces vary far more in the number of migrants they attract than in the number of migrants they send; destination effects vary more than origin effects. Nonetheless, half the variation in migration flows unexplained by the covariates cannot be attributed to origin and destination effects and instead relates to the unique interactions and relationships between pairs of provinces.

The estimated origin-destination correlation of 0.11 is small and not significant and so it is not the case that provinces that exhibit unusually high out-migration also exhibit unusually high or low in-migration. In contrast, the estimated flow-pair correlation of 0.72 is large and significant, suggesting that where one province sends a higher than predicted number of migrants to another, we in general also see a higher than predicted number of migrants sent from another province. Reciprocity in flows between provinces is clearly an



**Figure 2.** Predicted province origin random effects plotted in rank order with 95 per cent confidence intervals (unit: thousands).

important feature of urban–urban migration but this would have gone unnoticed in Model 2.

Table 3 presents the estimated correlations conditional on the covariates for the four forms of dependency, which further confirms findings of the random effects. Specifically, the model-implied correlation of flows sharing a common origin (Type 1) is 0.15, whilst that of flows sharing a common destination (Type 2) is 0.33 (more than twice of that of Type 1). However, the correlation between two residual flows where the destination of the first flow is the origin of the second (Type 3) to be just 0.02 and insignificant, and the correlation between reciprocal residual flows (Type 4) is 0.42.

Figure 2 plots the residual differences between the origins. They are shown in the original measurement units by exponentiating the predicted origin random effects and their 95 per cent confidence limits for each province, which is then compared to the reference line (Origin province effect = 1). Quantile-quantile plots (available on request) show that the predicted province origin (and destination) random effects are approximately normally distributed. Noticeably, the original reference line is ‘Origin province effect = 0’, which represents the theoretical mean of the normally distributed residuals. After exponentiating, the reference line of Fig. 2 becomes ‘Origin province effect = 1’, still representing the exponentiated theoretical mean but now being proportional to the overall average number of out-migrants across all provinces. The reason to do this is to make it more explicit to interpret the origin province effects on the original scale, as the data is log-transformed in the model. That is to say, the unit of the origin effects is thousand, as the

original data unit of urban migration is hundred and associated with a multiplier of 10 due to the 10 per cent sampling procedure. For instance, the origin random effect of Chongqing has the mean of 1.26 and an interval between 1 and 1.58 (Fig. 2), which does not overlap with the reference line and means that Chongqing significantly and systematically sends 260 more migrants on average than the overall national mean. In a similar way, based on whether the 95 per cent confidence intervals overlap the reference line or not, the provinces have been put into three groups. The first group contains provinces with above average residuals where the confidence intervals do not overlap with the overall average, indicating that they depart significantly from the theoretical gravity model by systematically exporting more urban–urban migrants than predicted by their population, income and distances to other provinces. The five provinces are Zhejiang, Fujian, Ningxia, Heilongjiang and Chongqing represented by the black dots. The second group contains provinces that are significantly below average and systematically export fewer migrants. They are Shanxi, Yunnan, Guangxi and Guizhou represented by the light grey dots. The remaining 22 provinces do not appear to have origin effects that deviate significantly from the overall average.

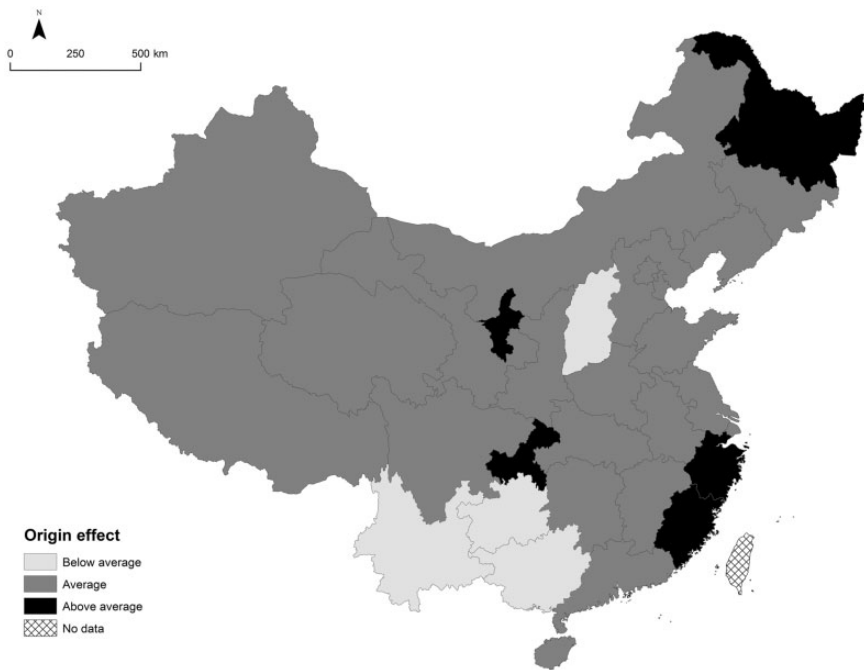
Figure 3 maps the residuals using the same colour scheme introduced in Fig. 2, indicating a strong and positive spatial autocorrelation overall, the spatial clustering of provinces with similar origin effects. For instance, three neighbouring provinces (Yunnan, Guizhou and Guangxi) with below-average exporting capabilities cluster at the south-western corner, two neighbouring coastal provinces (Fujian and Zhejiang) of above-average exporting abilities agglomerate in the southeast, whilst the majority of the provinces with average exporting capabilities form the biggest clustering in the map.

The destination effects can be considered in the same manner as the origin effects. In Fig. 4, there are nine provinces receiving significantly more migrants above the average (Xinjiang, Hainan, Guangdong, Beijing, Shaanxi, Ningxia, Gansu, Sichuan and Yunnan) and eight provinces that receive significantly less than average (Tibet, Henan, Anhui, Shanxi, Hunan, Tianjin, Jiangxi and Inner Mongolia). That the number of significant destination effects exceeds the number of significant origin effects is expected as destinations were shown to be twice as variable as the origins.

As shown in Fig. 5 mapping the destination effects reveals strong and positive spatial autocorrelation, but with more spatial variations than for the origin effects. In general, more spatial variations and more complicated spatial patterns can be observed in Fig. 5, with ribbon-like clusters and heterogeneous spots scattering all over the map.

## 5. Discussion

While rural–urban migration has remained important and has prevailed in developing countries, urban to urban migration has started to gain momentum as the developing world becomes more and more urbanised. This trend has already been observed in Latin American countries such as Brazil (Machado and Hakkert 1988), Mexico (Lozano-Ascencio et al. 1996), and Colombia (Shefer and Steinvortz 1993). China's urbanisation level has risen rapidly over the past three decades growing from 21 per cent in 1982 to 56 per cent in 2016. In 2010, around 90 million people migrated between urban areas in China.



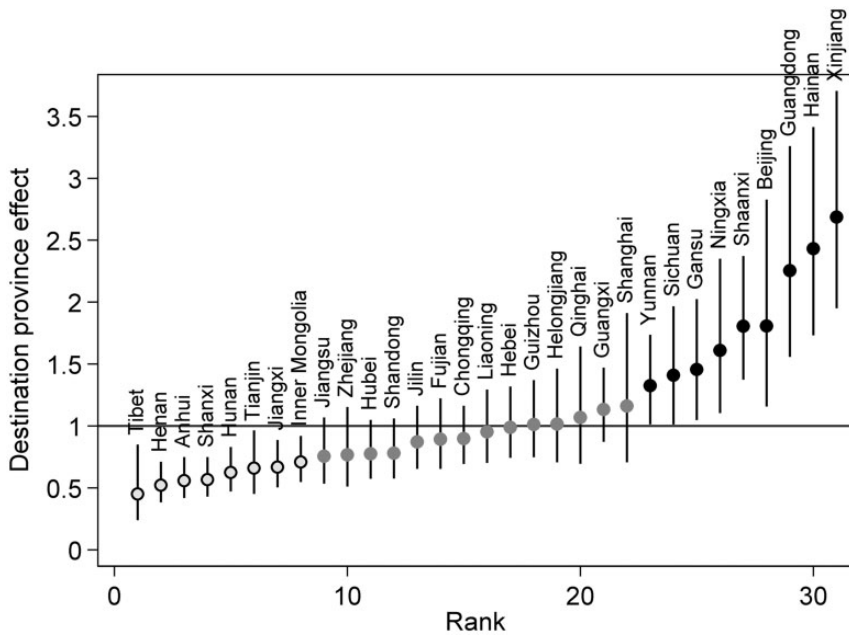
**Figure 3.** Spatial pattern of the predicted province origin random effects.

Identifying factors that affect inter urban migration flow helps shed light on the migratory process that urban–urban streams exhibit.

The results from Model 3 echo previous findings to some extent, namely the effect of origin, destination populations and incomes and distance on total migration (Fan 2005; Shen 2012; Liu et al. 2015). For urban populations, those of both origin and destination have significant positive effects upon urban–urban migration but that of the destination is overshadowed by its origin counterpart. Even though previous studies on total migration in China have shown similar results that the origin population is a more influential indicator for migration flows than the destination population (Fan 2005), this finding is a little counterintuitive for urban–urban migration. This perhaps is due to the fact that our study is using the urban population of the entire province instead of a particular city and therefore a larger urban population may not necessarily translate into the effect of a major urban labour market if the urban population is distributed across many medium and small size cities. Nevertheless, our models show that larger urban populations at both origin and destination provinces help to generate greater migration flows between them which agrees with existing findings relevant to migration stock theory (Fan 2005; Fan 2005; Shen 2015).

In terms of urban income, that of destination exerts significant positive impacts upon urban–urban migration, whereas origin urban income has strong negative effects by



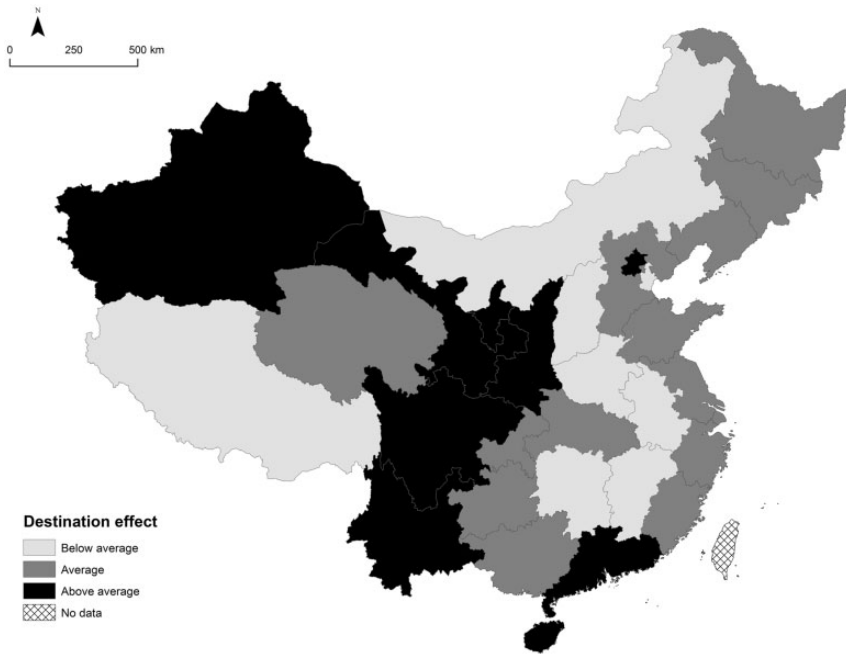


**Figure 4.** Predicted province destination random effects plotted in rank order with 95 per cent confidence intervals (unit: thousands).

contrast. This result is consistent with the conventional push-pull framework: migrants are pushed out of areas with lower income and attracted to areas with higher earning. The gap between the effect sizes of origin and destination urban income signposts that the economic pull force plays a much bigger role than push factors in this migration system which agrees with the general observations that pull factors outweigh push factors in most migration flows, particularly for economic migration. This also suggests that the urban–urban migration in China is mainly economically driven.

With respect to distance, it has a substantial negative effect upon urban–urban migration flows: as the distance between the provincial capital cities increases by 10 per cent, the migration stream reduces by 11 per cent. This is surprising given the substantial investment China has made in domestic infrastructure and transportation and the associated reductions in travelling cost which this has brought about in recent decades (Luo, Zhu and Zou 2014). Still, the deterrence effect of distance on urban–urban migration flow is consistent with existing findings (Yan 2007; Shen 2013).

What the multilevel model shows, which standard approaches cannot, is that all the random effect parameters, except for the origin–destination correlation coefficient, are significant and contribute substantially to explain the interprovincial urban–urban migration flow residuals. Adding origin and destination variances, as well as allowing flow-pairs to correlate, greatly improves the fit of the model.



**Figure 5.** Spatial pattern of the predicted province destination random effects.

Residual differences in the number of migrants leaving the provinces seem to be closely linked with the provinces' urbanisation level. Except for Ningxia, provinces with significantly higher than average exporting capabilities all had an above the national average urbanisation level (50 per cent) in 2010 ranging from 62 per cent in Zhejiang to 53 per cent in Chongqing. Even Ningxia's urbanisation level was only slightly lower than the average with 48 per cent. Provinces with significantly lower exporting capabilities all had below average urbanisation level. In fact, other than Shanxi (48 per cent), the remaining provinces were among the least urbanised provinces in China. Guizhou, Yunnan, Gansu and Guangxi were ranked near the bottom in terms of urbanisation, and are all located in less developed Western China. This finding lends support to a possible hypothesis raised in the previous section that even though origin urban population helps increase migration flow, more urbanised provinces may provide more opportunities and therefore do not have such a strong push force as less urbanised provinces.

The destination effects are more variable. For provinces that deviate significantly from the average, it appears that there are different patterns, some of which correspond to those of internal migration and others that do not. For provinces that exhibit higher than average attraction, Beijing and Guangdong have known to be top migration destinations over the past three decades due to their high level of economic development. The rest of the provinces are mostly in Western China except for Hainan in the south. Among them, Xinjiang

has also been a major migration destination due to its abundant natural and land resources and policy-led development (Fan 2005; Liu et al. 2014). Neither Ningxia nor Hainan receive large volume of migrants but appear to have the top ten in-migration rate between 2005 and 2010 in the nation, which is related to the comparatively smaller population size (Liu et al. 2014). The rest of the Western provinces are more known for sending migrants instead of receiving. Another interesting fact about those provinces is that there is a distinctive pattern in terms of their urban migration rate. The three coastal provinces (Beijing, Guangdong and Hainan) all have a fairly high urban migration rate ranging from 0.91 in Hainan to 1.69 in Beijing (ranked 1st) and the Western provinces all have lower urban migration rate from 0.41 in Yunnan to 0.61 in Gansu. The above results suggest that even though the Western provinces are less developed and therefore not as active in stimulating urban–urban migration flows, these Western provinces exert higher than average attraction to urban migrants by offering other possibilities such as higher potential of income growth. This may be explained by the recent higher than average economic growth Western provinces have experienced due to the increased investment and preferential policies as part of China Western Development program sponsored by the central government.

In terms of provinces with below average attractiveness, except for Tianjin all are in less developed non-coastal areas with the majority located in central China. Even though Tianjin is a very developed municipality, its geographical proximity to Beijing perhaps explains why urban migration flows respond passively to its urban labour market (population) as Beijing is a more popular destination for urban migrants.

At the flow-pair level, an interesting fact is that among the flows that exhibit the largest and smallest residuals, the majority of them are between Western provinces followed by those involving a Western province as an origin or destination. Western provinces are the least developed area in China that has experienced fast policy-led development recently. Their economic growth could either act as a pull force due to increased economic opportunities or a push force since the development would enable those who were not able to migrate to move to more developed areas now. Those facts might help contribute to their irregular or unpredictable urban migration flows from the more established migration streams. In addition, we also find high correlations between some of the pair-flows and most of those tend to be between Western provinces as well. For instance, the reciprocal migration flows between Yunnan and Xinjiang which are distantly apart are both significantly lower than predicted, while the bilateral migration flows between neighbouring provinces Sichuan and Tibet are both significantly higher than predicted. The latter result in a way mirrors international migration in less developed world. For example, most migrants in Africa and Southeast Asia tend to go to nearby countries instead of countries further away with more job opportunities. Only when countries become more developed, more people would start to engage in longer distanced migration.

## 6. Conclusions

This paper has set out a method of multilevel modelling better to understand the patterns of interprovincial urban–urban migration in China. Whilst prior studies have greatly contributed to the understanding of migration flows, they wrongly treat migration flows

between shared origins and/or destinations as independent events, leading to potentially inaccurate results, as well as overlooking substantively interesting correlations in the patterns of movement. In addition, little attention has been given to urban–urban migration flows. Our endeavour has overcome both research gaps, which are successfully considered in this paper.

This paper is the first to systematically analyse urban–urban migration in China, a phenomenon which is on the rise in developing countries that have been going through rapid urbanisation. Comparing to internal migration in China which is dominated by rural–urban migration, urban–urban migration is similar in the sense that it is also economically driven and larger population sizes at both origins and destinations help to contribute to the volume of the migration flows. Moreover, distance plays a sizeable deterring role on urban–urban migration, which may have important policy implications. On the one hand, for major cities in the western and inland areas, designing policies to manage distance's adverse effect can help reducing regional inequality by encouraging information, skill and capital transfer carried out by urban–urban migrants between them and the coastal cities. On the other hand, as China has been interested in initiating localised urbanisation by favouring developing small- and medium-sized cities since the 1980s (Han and Yan 1999; Chen, Liu, and Lu 2013), the distance decay effect may reduce long distance urban–urban migration and as a result facilitate the local socio-economic development which will greatly benefit smaller cities.

The findings also suggest that development level is closely linked with urban–urban migration. For example, urbanisation level plays a key role in provinces' origin effect. The destination effect and pair-flow effects also indicate how the policy-driven growth in the least developed Western region has had great influence on urban–urban migration in China.

## Funding

This work was supported by 'Economic and Social Research Council' (Grant number: S133721-119).

*Conflict of interest statement.* None declared.

## References

- Andersson, A. E. et al. (2012) *Interregional Migration: Dynamic Theory and Comparative Analysis*. Salmon Tower Building, New York City: Springer Science & Business Media.
- Beine, M., Bertoli, S. and Moraga, J. F.-H. (2014) 'A Practitioners' Guide to Gravity Models of International Migration', *Documentos de trabajo (FEDEA)*, 1–27.
- Berry, B. J. L. (1976) *Urbanization and Counter-Urbanization*. Thousand Oaks, CA: SAGE.
- Brandt, L. and Holz, C. A. (2006) 'Spatial Price Differences in China: Estimates and Implications', *Economic Development and Cultural Change*, 55/1: 43–86.
- Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2011) 'Robust Inference with Multiway Clustering', *Journal of Business & Economic Statistics*, 29/2: 238–49.

- Cao, G. Z. et al. (2017) 'Consumption Structure of Migrant Worker Families in China', *China & World Economy*, 25/4: 1–21.
- Cerrutti, M. and Bertoncello, R. (2003) *Urbanization and Internal Migration Patterns in Latin America*. Buenos Aires, Argentina: Centro de Estudios de Población.
- Chen, Z., Liu, X. F. and Lu, M. (2013) 'Beyond Lewis: Rural-to-Urban Migration with Endogenous Policy Change', *China Agricultural Economic Review*, 5/2: 213–30.
- Cheng, Z. M., Nielsen, I. and Smyth, R. (2014) 'Access to Social Insurance in Urban China: A Comparative Study of Rural–Urban and Urban–Urban Migrants in Beijing', *Habitat International*, 41: 243–52.
- Christian, W. and Braden, W. (1966) 'Rural Migration and Gravity Model', *Rural Sociology*, 31: 73–80.
- Claeson, C. F. (1969) '2-Stage Model of in-Migration to Urban Centres—Deductive Development of a Variant of Gravity Formulation', *Geografiska Annaler Series B-Human Geography*, 51/2: 127–38.
- Converse, P. D. (1949) 'New Laws of Retail Gravitation', *Journal of Marketing*, 14/3: 379–84.
- De Haas, H. (2010a) 'The Internal Dynamics of Migration Processes: A Theoretical Inquiry', *Journal of Ethnic and Migration Studies*, 36/10: 1587–617.
- (2010b) 'Migration and Development: A Theoretical Perspective', *International Migration Review*, 44/1: 227–64.
- Dennett, A. and Wilson, A. (2013) 'A Multilevel Spatial Interaction Modelling Framework for Estimating Interregional Migration in Europe', *Environment and Planning A*, 45/6: 1491–507.
- Ezra, M. and Kiros, G. E. (2001) 'Rural Out-Migration in the Drought Prone Areas of Ethiopia: A Multilevel Analysis', *International Migration Review*, 35/3: 749–71.
- Fan, C. C. (2005) 'Interprovincial Migration, Population Redistribution, and Regional Development in China: 1990 and 2000 Census Comparisons', *The Professional Geographer*, 57/2: 295–311.
- (2005) 'Modeling Interprovincial Migration in China, 1985–2000', *Eurasian Geography and Economics*, 46/3: 165–84.
- and Wang, W. W. (2008) 'The Household as Security: Strategies of Rural–Urban Migrants in China', *Migration and Social Protection in China*, 14: 205.
- Feng, Y., Kugler, J. and Zak, P. J. (2002) 'Population Growth, Urbanisation and the Role of Government in China: A Political Economic Model of Demographic Change', *Urban Studies*, 39/12: 2329–43.
- Fingleton, B. (1986) 'Analyzing Cross-Classified Data with Inherent Spatial Dependence', *Geographical Analysis*, 18/1: 48–61.
- Flowerdew, R. and Aitkin, M. (1982) 'A Method of Fitting the Gravity Model Based on the Poisson Distribution', *Journal of Regional Science*, 22/2: 191–202.
- and Salt, J. (1979) 'Migration between Labour Market Areas in Great Britain, 1970–1971', *Regional Studies*, 13/2: 211.
- Getis, A. (1990) 'Screening for Spatial Dependence in Regression-Analysis', *Papers of the Regional Science Association*, 69/1: 69–81.

- Giles, J. and Yoo, K. (2007) 'Precautionary Behavior, Migrant Networks, and Household Consumption Decisions: An Empirical Analysis Using Household Panel Data from Rural China', *Review of Economics and Statistics*, 89/3: 534–51.
- Ginsberg, R. B. (1972) 'Incorporating Causal Structure and Exogenous Information with Probabilistic Models—With Special Reference to Choice, Gravity, Migration, and Markov Chains', *Journal of Mathematical Sociology*, 2/1: 83–103.
- Goldstein, H. (2011) *Multilevel Statistical Models*. Hoboken, NJ: John Wiley & Sons.
- Gordon, I. (1995) 'Migration in a Segmented Labour Market', *Transactions (Institute of British Geographers)*, 20/2: 139–55.
- Greenwood, M. J. and Hunt, G. L. (1989) 'Jobs Versus Amenities in the Analysis of Metropolitan Migration', *Journal of Urban Economics*, 25/1: 1–16.
- Hahn, H. P. (2010) 'Urban Life-Worlds in Motion: In Africa and Beyond', *Africa Spectrum*, 45: 115–29.
- Han, X. R. and Chen, Y. F. (2016) 'Food Consumption of Outgoing Rural Migrant Workers in Urban Area of China a QUAIDS Approach', *China Agricultural Economic Review*, 8/2: 230–49.
- Han, S. S. and Yan, Z. X. (1999) 'China's Coastal Cities: Development, Planning and Challenges', *Habitat International*, 23/2: 217–29.
- Harris, R. (2017) 'Fitting a Multilevel Index of Segregation in R', in *Using the MLID Package*. <[https://www.researchgate.net/profile/Richard\\_Harris5/publication/313903303\\_Fitting\\_a\\_multilevel\\_index\\_of\\_segregation\\_in\\_R\\_using\\_the\\_MLID\\_package/links/58af24c392851cf7ae88f5ed/Fitting-a-multilevel-index-of-segregation-in-R-using-the-MLID-package.pdf](https://www.researchgate.net/profile/Richard_Harris5/publication/313903303_Fitting_a_multilevel_index_of_segregation_in_R_using_the_MLID_package/links/58af24c392851cf7ae88f5ed/Fitting-a-multilevel-index-of-segregation-in-R-using-the-MLID-package.pdf)> accessed 23 July 2018.
- Harris, J. R. and Todaro, M. P. (1970) 'Migration, Unemployment and Development: A Two-Sector Analysis', *The American Economic Review*, 60: 126–42.
- Jiang, Y. et al. (2017) 'Rural Migrant Workers' Intentions to Permanently Reside in Cities and Future Energy Consumption Preference in the Changing Context of Urban China', *Transportation Research Part D-Transport and Environment*, 52: 600–18.
- Johnston, R. et al. (2016) 'House Price Increases and Higher Density Housing Occupation: The Response of Non-White Households in London, 2001–2011', *International Journal of Housing Policy*, 16/3: 357–75.
- Kallan, J. E. (1993) 'A Multilevel Analysis of Elderly Migration', *Social Science Quarterly*, 74: 403–16.
- Kenny, D. A. and Kashy, D. A. (2011) 'Dyadic Data Analysis Using Multilevel Modeling', in *Handbook of Advanced Multilevel Analysis*, pp. 344–60.
- King, R. (2012) 'Theories and Typologies of Migration: An Overview and a Primer'. <<https://www.mah.se/upload/Forskningscentrum/MIM/WB/WB%203.12.pdf>> accessed 23 July 2018.
- Koster, J. M. and Leckie, G. (2014) 'Food Sharing Networks in lowland Nicaragua: An Application of the Social Relations Model to Count Data', *Social Networks*, 38: 100–10.
- Laczko, F., Tjaden, J. and Auer, D. (2017) 'Measuring Global Migration Potential, 2010–2015', Global Migration Data Briefing Series. <[https://publications.iom.int/system/files/pdf/gmdac\\_data\\_briefing\\_series\\_issue\\_9.pdf](https://publications.iom.int/system/files/pdf/gmdac_data_briefing_series_issue_9.pdf)> accessed 23 July 2018.
- Leamer, E. E. (2012) *The Craft of Economics: Lessons from the Heckscher-Ohlin Framework*. Cambridge, MA: MIT Press.

- Leckie, G. (2013) 'Cross-Classified Multilevel Models-Concepts'. *LEMMA VLE Module*, 12: 1–60.
- and Charlton, C. (2013) 'Runmlwin—A Program to Run the MLwiN Multilevel Modelling Software from Within Stata', *Journal of Statistical Software*, 52/11: 1–40.
- Leorato, S. and Mezzetti, M. (2016) 'Spatial Panel Data Model with Error Dependence: A Bayesian Separable Covariance Approach', *Bayesian Analysis*, 11/4: 1035–69.
- Lewis, W. A. (1954) 'Economic Development with Unlimited Supplies of Labour', *The Manchester School*, 22/2: 139–91.
- Liu, Y. et al. (2014) 'Interprovincial Migration, Regional Development and State Policy in China, 1985–2010', *Applied Spatial Analysis and Policy*, 7/1: 47–70.
- Liu, T. et al. (2015) 'Spatial Patterns, Driving Forces, and Urbanization Effects of China's Internal Migration: County-Level Analysis Based on the 2000 and 2010 Censuses', *Journal of Geographical Sciences*, 25/2: 236–56.
- Lozano-Ascencio, F., Roberts, B. R. and Bean, F. D. and Population Research Center, University of Texas at Austin (1996) *The Interconnectedness of Internal and International Migration: The Case of the United States and Mexico*. Texas Population Research Center, The University of Texas at Austin.
- Luo, X. B., Zhu, N. and Zou, H. F. (2014) 'China's Lagging Region Development and Targeted Transportation Infrastructure Investments', *Annals of Economics and Finance*, 15: 365–409.
- Machado, C. C. and Hakkert, R. (1988) '[An Exploratory Analysis of Information on Inter- and Intra-Municipal Migration in the 1980 Census]', *Rev Bras Estud Popul*, 5: 1–20.
- Massey, D. S. et al. (1993) 'Theories of International Migration: A Review and Appraisal', *Population and Development Review*, 19/3: 431–66.
- Molho, I. (2013) 'Theories of Migration: A Review', *Scottish Journal of Political Economy*, 60/5: 526–56.
- Poot, J. (1986) 'A System Approach to Modelling the Inter-Urban Exchange of Workers in New Zealand', *Scottish Journal of Political Economy*, 33/3: 249–74.
- Pumain, D. and Sanders, L. (2013) 'Theoretical Principles in Interurban Simulation Models: A Comparison', *Environment and Planning A*, 45/9: 2243–60.
- Rasbash, J. et al. (2009) 'MLwiN Version 2.26'. Centre for Multilevel Modelling, University of Bristol. <<http://www.bris.ac.uk/media-library/sites/cmm/migrated/documents/manual-web.pdf>> accessed 23 July 2018.
- Remoundou, K., Gkartzios, M. and Garrod, G. (2016) 'Conceptualizing Mobility in Times of Crisis: Towards Crisis-Led Counterurbanization?' *Regional Studies*, 50/10: 1663–74.
- Shefer, D. and Steinvortz, L. (1993) 'Rural-to-Urban and Urban-to-Urban Migration Patterns in Colombia', *Habitat International*, 17/1: 133–50.
- Shen, J. F. (2012) 'Changing Patterns and Determinants of Interprovincial Migration in China 1985–2000', *Population, Space and Place*, 18/3: 384–402.
- (2013) 'Increasing Internal Migration in China from 1985 to 2005: Institutional Versus Economic Drivers', *Habitat International*, 39: 1–7.
- (2015) 'Explaining Interregional Migration Changes in China, 1985–2000, Using a Decomposition Approach', *Regional Studies*, 49/7: 1176–92.
- (2016) 'Error Analysis of Regional Migration Modeling', *Annals of the American Association of Geographers*, 1–15.



- Taylor, E. J. (1999) 'The New Economics of Labour Migration and the Role of Remittances in the Migration Process', *International Migration*, 37/1: 63–88.
- Thomas, M., Stillwell, J. and Gould, M. (2015) 'Modelling Multilevel Variations in Distance Moved Between Origins and Destinations in England and Wales', *Environment and Planning A*, 47/4: 996–1014.
- Todaro Michael, P. (1976) *Internal Migration in Developing Countries: A Review of Theory, Evidence, Methodology and Research Priorities*. Geneva: International Labour Office.
- Vartiainen, P. (1989) 'Counterurbanisation: A Challenge for Socio-Theoretical Geography', *Journal of Rural Studies*, 5/3: 217–25.
- Vignoli, J. R. (2008) 'Spatial Distribution, Internal Migration and Development in Latin America and the Caribbean', *Cepal Review No. 96*, 137–57.
- Wilson, A. G. (1971) 'A Family of Spatial Interaction Models, and Associated Developments', *Environment and Planning A: Economy and Space*, 3/1: 1–32.
- Wu, Z. M. and Yao, S. J. (2003) 'Intermigration and Intramigration in China: A Theoretical and Empirical Analysis', *China Economic Review*, 14/4: 371–85.
- Yan, S. (2007) 'Inter-Provincial Migration and Its Determinants in 1990's China', *Chinese Journal of Population Science*, 1: 71–7.
- Yang, J. H. (2013) 'Social Exclusion and Young Rural–Urban Migrants' Integration into a Host Society in China', *Annals of the American Academy of Political and Social Science*, 648/1: 52–69.
- Yang, X. S. and Guo, F. (1999) 'Gender Differences in Determinants of Temporary Labor Migration in China: A Multilevel Analysis', *International Migration Review*, 33/4: 929–53.
- Zarate-Hoyos, G. A. (2004) 'Consumption and Remittances in Migrant Households: Toward a Productive Use of Remittances', *Contemporary Economic Policy*, 22/4: 555–65.
- Zukin, S. (1998) 'Urban Lifestyles: Diversity and Standardisation in Spaces of Consumption', *Urban Studies*, 35: 825–39.

## Appendix

This appendix presents the derivations for the four model implied correlations (Type 1, 2, 3 and 4) presented in Table 1. Reconsider Model 3 (Equation 5), the most general model presented in Section 3. The conditional variance for migration flow  $y_{ij}$  is given by

$$\text{Var}(y_{ij} \mid \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}) = \text{Var}(o_i + d_j + e_{ij}) = \sigma_o^2 + \sigma_d^2 + \sigma_e^2$$

while the conditional covariance between two migration flows  $y_{ij}$  and  $y_{i'j'}$  is given by

$$\text{Cov}(y_{ij}, y_{i'j'} \mid \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1i'}, \mathbf{x}_{2j'}, \mathbf{x}_{3i'j'}) = \text{Cov}(o_i + d_j + e_{ij}, o_{i'} + d_{j'} + e_{i'j'})$$

and will vary in strength depending on the extent to which the origin and destination provinces of each flow overlap or not. The conditional correlation can then be calculated in the usual way as follows



$$\text{Corr}\left(y_{ij}, y_{i'j'} \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1i'}, \mathbf{x}_{2j'}, \mathbf{x}_{3i'j'}\right) = \frac{\text{Cov}\left(o_i + d_j + e_{ij}, o_{i'} + d_{j'} + e_{i'j'}\right)}{\sqrt{\text{Var}\left(o_i + d_j + e_{ij}\right)} \sqrt{\text{Var}\left(o_{i'} + d_{j'} + e_{i'j'}\right)}}$$

The four types of correlation presented in Table 1 can then be derived as follows:

Type 1: The correlation between the flow from origin  $i$  to destination  $j$  and from origin  $i$  to destination  $k$  is given by

$$\text{Corr}(y_{ij}, y_{ik} \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1i}, \mathbf{x}_{2k}, \mathbf{x}_{3ik}) = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$$

Type 2: The correlation between the flow from origin  $i$  to destination  $j$  and from origin  $j$  to destination  $k$  is given by

$$\text{Corr}(y_{ij}, y_{jk} \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1j}, \mathbf{x}_{2k}, \mathbf{x}_{3jk}) = \frac{\sigma_{od}}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$$

Type 3: The correlation between the flow from origin  $i$  to destination  $j$  and from origin  $k$  to destination  $j$  is given by

$$\text{Corr}(y_{ij}, y_{kj} \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1k}, \mathbf{x}_{2j}, \mathbf{x}_{3kj}) = \frac{\sigma_d^2}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$$

Type 4: The correlation between the flow from origin  $i$  to destination  $j$  and from origin  $j$  to destination  $i$  is given by

$$\text{Corr}(y_{ij}, y_{ji} \mathbf{x}_{1i}, \mathbf{x}_{2j}, \mathbf{x}_{3ij}, \mathbf{x}_{1j}, \mathbf{x}_{2i}, \mathbf{x}_{3ji}) = \frac{2\sigma_{od} + \sigma_{ee}}{\sigma_o^2 + \sigma_d^2 + \sigma_e^2}$$

Model 1 (Equation 3) is a constrained version of Model 3 where  $\sigma_o^2 = 0$ ,  $\sigma_d^2 = 0$ ,  $\sigma_{od} = 0$  and  $\sigma_{ee} = 0$  and so all four correlations are implicitly assumed to be zero. Model 2 (Equation 4) is a constrained version of Model 3 where  $\sigma_{od} = 0$  and  $\sigma_{ee} = 0$  and so the last two correlations are implicitly assumed to be zero.